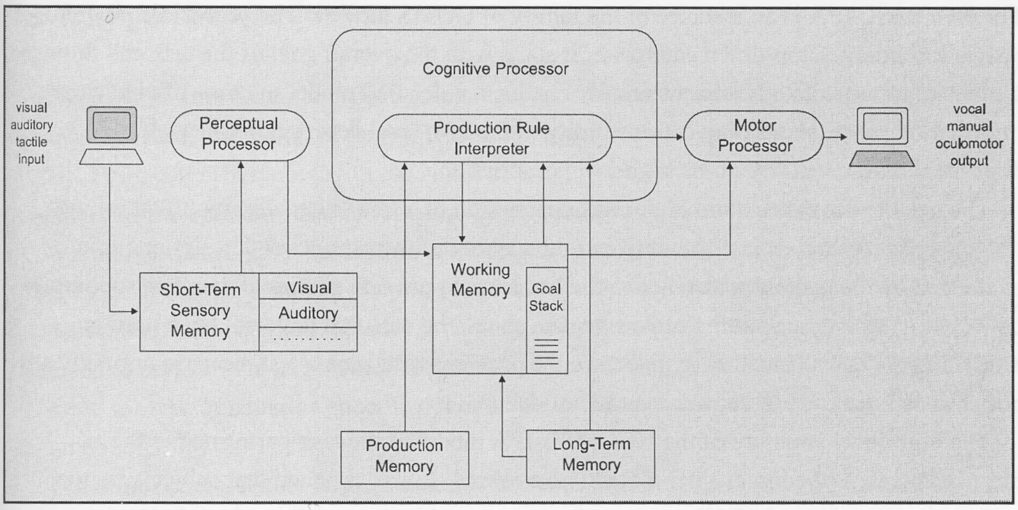
Cognitive task analysis in human-computer interaction

Lamain, W.

Chapter 8:
General discussion and conclusions



Within this chapter the conclusions from the previous chapters are summarized, integrated and discussed. It is concluded that the NGOMSL-IPA is a valid and usable cognitive task analysis method that is especially suited to model information processing and mental workload. Also it is concluded that the methodology used in this thesis is usable for the inclusion of psychophysiology in human-computer interaction research. Finally, the necessity of sound experimentation is assessed. Also some useful extensions to the NGOMSL-IPA are suggested as well as suggestions for a better use of both the NGOMSL-IPA method and psychophysiology within human-computer interaction research.

8.1 The NGOMSL-IPA approach

The NGOMSL-IPA approach was put forward as the approach to meet the requirements for a cognitive task analysis that were discussed in the introduction. The information processing modelling was the first major requirement. In addition, it should enable an estimation of the mental workload involved in task performance. The approach should be a predictive one. Finally, it also should provide an overview of the task structure and a time-line analysis. The approach is based on the NGOMSL approach of Kieras (1991; 1993; 1996; 1999), and the original approach already met some of these requirements. Especially the requirements for the estimation of mental workload and information processing profile, required adaptations. In the following paragraphs it will be evaluated in how far the requirements were met.

8.1.1 Task structure and NGOMSL-IPA

The NGOMSL-IPA is an instance of the family of GOMS models. The process of modelling a task is essentially a top-down enterprise. It starts with the overall goal of the task and through a series of intermediate levels, eventually results in a detailed model in terms of elementary information processing actions. The validity of the low-level description of a task is dependent on the validity of the high-level description.

The need to provide a valid high-level description of a task is equally important for any task analysis method. It is difficult to test the validity of the high-level description, as is evidenced by the general reluctance in the literature to provide empirical data that support (or not support) task description. Supportive data should be data that have not been used in modelling the task. This directly indicates the problem with such a test, because normally all possible data are used to construct a task model.

The high-level structure of the NGOMSL-IPA model of the task performed in the experiments described in chapter 6 could be assessed. It was assumed that subjects try to achieve the overall goal of the task by subsequently achieving smaller subgoals (Anderson, 1993). This requires a goal-stack like structure to control behavior. Popping subgoals from the stack and pushing subgoals on the stack takes time. The intervals between keypresses that belong to different subgoals are prolonged because they contain 'push' and 'pop' operators. The actual keypress data showed that length of the inter-keypress intervals varied with the number of goal stack operators they contained (according to the NGOMSL-IPA task model).

The fact that the length of the intervals varied with the number of goal stack operators evidences the validity of the model. Subjects decomposed the task exactly as it was decomposed in the task model, pausing at the moments when subtasks were completed and new subgoals had to be set. Thus the method described in chapter 2, which is grounded in psychological theory and empirical evidence, correctly captures the essential structure of a task. When properly applied, the method results in a valid task model.

Task performance is vulnerable to disturbances at the transitions between sub-goals, especially when substantial changes in the goal stack are executed. If possible, at these moments the task performer should not be interrupted or loaded with additional information. The fact that the NGOMSL-IPA model can identify error-prone moments in task performance and moments at which high level behavior planning occurs, is of large importance for designing tasks and artefacts.

The inter-keypress data in relation to the goal stack operators also prove that subjects seem to make use of a goal stack like structure while performing a task. This had already been shown by Anderson (1993), Anderson & Lebiere (1998), Egan & Greeno (1974) and Ruiz (1987) for problem solving tasks. The data from chapter 6 indicate that also in more routine cognitive skill tasks a goal stack is used.

8.1.2 Extensions of the NGOMSL approach

8.1.2.1 Specific operators

The NGOMSL-IPA (Natural GOMS Language-Information Processing Analysis) approach is an adaptation and extension of the NGOMSL approach from Kieras (1991; 1993; 1996; 1999). The exact differences with the NGOMSL approach are discussed in chapter 2. Here, a summary of the operators that differ from the original NGOMSL will be given.

Within NGOMSL-IPA, perception is modelled by only two operators, which are different from those of Kieras. Simple perception is performed by the 'Perceive <item>' operator, and perception of more complex material is performed by the 'Read <item>' operator. These are believed to perform the largest part of perception (not reading a sentence or story) and prevent the need for task or situation specific operators as in the approach of Kieras (e.g. 'Locate menu option').

The motor operators are largely identical between the NGOMSL and NGOMSL-IPA approach, except for the operators that perform mouse actions. Kieras uses the press key operator to click the mouse. In chapter 4 it was argued that three operators were needed and sufficient to operate the mouse: 'Press <left/middle/right> mouse button', 'Release <left/middle/right> mouse button', and 'Click <left/middle/right> mouse button'. Scrolling with the mouse was not included.

The use of working memory and long term memory differs markedly from the approach of Kieras. Kieras uses a specific 'Forget from WM' operator to remove information from working memory. Within the NGOMSL-IPA there is not such an operator, although it must be indicated when information can be forgotten, it just is not modelled as a deliberate act.

Long term memory is seen as a network of associated items, which must be activated to a certain level to be retrieved. Association spreads from one item to related items, and decays over time (Anderson, 1983). It was recognized that these characteristics of human memory have large implications for the speed of memory retrieval and the load on memory in human-computer interaction. The operator that performs retrieval from long term memory is extended with two parameters, one indicating the frequency of retrieval and one indicating the degree of recall or recognition: 'Retrieve from LTM that <LTM-object-description> [recall-frequency]'.

The differences between recall and partly recall/partly recognition was evidenced in chapter 7. Recall proved to be slower than partly recall/partly recognition. The information processing profile distinction between the conditions were for a large part based on the differences between recall and partly recall/partly recognition memory operators. Consequently there was a clear difference between the conditions in terms of load on the cognitive processor. This evidences the differences in cognitive load the different memory operators induce. The frequency of retrieval parameter also proved to be valid. Facts that were frequently retrieved were retrieved faster than facts that were retrieved less frequently.

8.1.2.2 Estimating mental workload

An important impetus for the development of the NGOMSL-IPA approach was the one-sided emphasis on task performance, within existing GOMS-like methods. Time to perform a task and time needed to learn a task are examples of this emphasis. It was argued in chapter 1 that in addition to these performance aspects, also the workload involved in performing tasks is essential information. The NGOMSL-IPA is a first attempt to further elucidate this aspect. Several estimates of mental workload and information processing load have been tested and they turned out to be valid, i.e. an NGOMSL-IPA model can be used to predict both the costs involved in task performance and the processing underlying task performance. This information can be used, in addition to performance time, learning time and cognitive complexity estimates, in designing human-computer interfaces that optimally comply with human capacities and skills.

There is one straight forward way to include mental workload, simply on the basis of the standard time estimation. As Mulder, Mulder and Veldman (1985) proposed, time pressure can be considered a major contributor to mental workload. This is also evidenced by the inclusion of a time pressure dimension in the NASA-TLX workload scale (Hart & Staveland, 1988). Mulder et al. proposed a simple index of mental workload by dividing the time needed to perform a task by the time available to perform a task (an approach that is also adopted by Neerincx (1995; 1999). A quotient of exactly one, indicates that there is just enough time to perform the task, a quotient below one means that there is more time than needed, while a quotient higher than one indicates that the time is not sufficient for adequate task performance. The latter situation indicates a high workload situation (Parks & Boucek, 1989). The time needed to perform a task can easily be calculated by adding the times of the individual operators. The time available cannot always easily be assessed, but usually in a working environment, the total amount of work to be performed can be ascertained and can be used to calculate the total amount of time available for one task.

Mental workload is multi-dimensional (O'Donnel & Eggemeier, 1986; Hart & Staveland, 1988). The time needed divided by time available index, only taps one dimension of workload. Another index that more directly relates to the information processing requirements of a task, is the load on working memory. An NGOMSL-IPA model specifies when information is retained in working memory, and also when it can be forgotten. Therefore a continuous count of the number of chunks in working memory can be made (see Kieras, 1993;

Lerch, Mantei & Olson, 1989). Working memory load should be expressed in two measures: an average load and a peak load. The average load is the average number of chunks in working memory, and the peak load is the highest number of chunks in working memory at some moment during task performance. Generally speaking, the more chunks that have to be retained in working memory, the higher the mental workload. In addition, a peak load that approximates or exceeds that maximal capacity of working memory (5-9 chunks), incurs a very high mental workload (Card, Moran & Newell, 1983).

A third index of mental workload that can be calculated from an NGOMSL-IPA task model is the depth of a goal structure. It was argued that as subjects work deeper in a goal structure, i.e. if there are more superordinate goals, then the load on working memory is also higher. They must keep track of the higher level goals in order to resume their attainment as soon as the current goal is achieved. Depth of the goal structure in an NGOMSL-IPA task model is analogous with height of the goal stack in working memory. It could be that the goal stack resides in normal working memory, and thus occupies some of the 7±2 chunks of capacity. If it does, it is somewhat strange that that has never been discovered in the many working memory capacity experiments, although these have never been designed to specifically test the effects of the height of a goal stack on regular working memory capacity. On the other hand, it could very well be that there is a special structure in working memory that is dedicated to storing the goal stack, analogous to the visuo-spatial scratch pad for the storage of specific visual-spatial information (Baddeley & Hitch, 1974; Gathercole, 1994). The central executive structure is believed to be involved in controlling working memory and the conscious control of behavior and is believed to control the goal stack. Further research is needed to clarify this issue.

The two text-editing experiments in chapter 6 showed that behavior as well as psychophysiological measures were sensitive to the depth of the goal structure. Performance was worse when subjects were performing a part of a task that was located deep in a goal structure, while in addition psychophysiological measures indicated that workload was higher for subtasks located deeper in the goal structure.

All features of an NGOMSL-IPA task model that are related to mental workload and that were mentioned above, are well suited for making a relative estimation of mental workload between different tasks. It is not possible to make an absolute estimation of workload, i.e. one cannot express workload in an absolute number which can be regarded situation-independent, such as performance time can be. At best, workload estimation can be used as a relative index with which several tasks can be compared.

A second problem is that it is difficult to state when mental workload is too high or too low. An overload clearly is not desirable. If there is less time available for performing a task than is needed, and the average working memory load is close to the upper limit (7-9 chunks), and the peak load on working memory is more than the upper limit, and the goal structure is very deep (6 levels or more), then it is clear that the task will incur a very high workload. On the other hand, if there is much more time than needed, working memory load is low, and peak load also, and the goal structure is only a few levels deep, then the task will not incur much workload. The problem is that usually tasks will be located somewhere in between these extremes. In such cases it will be difficult to say what the most optimal workload will be. As yet, choosing the task with the most optimal workload from several alternatives, is dependent on the skills, knowledge and intuition of the task analyst.

In case serious problems have been reported by task performers, then the use of the mental workload estimates are somewhat more straight forward. Then they can be used to diagnose the problem in the task. More generally speaking, if data from other sources, such as behavior or interviews, indicate that there is a problem with a task, then the mental workload estimates can be very useful in checking whether there is a problem with workload. In addition, specific bottlenecks in task performance can be predicted, based on the mental workload estimates. Consequently, it can be used to analyse these specific task parts in more detail.

The best solution to the problem of the value of the actual level of workload, probably is to design a data base of tasks that can be used as a reference table. Any task could then be compared to the tasks in the data base and its workload be expressed relative to other tasks and other groups of the working population. Another possibility is to design a number of standard tasks and express mental workload relative to the workload involved with performing those standard tasks. These are possible future developments that would be needed if the issue of mental workload is to be of practical use in human-computer interaction.

8.1.2.3 Information processing profile

Another main extension of the NGOMSL approach that was achieved in the NGOMSL-IPA approach is the inclusion of an information processing profile. The validity of such a profile was shown in chapter 7.

The actual use of an information processing profile in human-computer interaction is still somewhat vague. Most important is that it forces the analyst to think of a task in terms of the information processing mechanisms underlying performance. This way the analyst will be forced to use a different frame of reference and will be forced to emphasize other aspects of task performance. This is a major change from many traditional task analysis techniques that only relate to observable behavior. Especially in computer tasks, that are predominantly cognitive in character, the non-observable cognitive behavior is essential, and calls for a cognitive task analysis approach.

The actual summary of information processing as is given by the information processing profile, characterizes the average information processing requirements of the task. It can reveal subtle differences that may not seem very profound for the individual task parts, but clearly show in the overall task.

Also, differences between tasks, that generally seem to be overlooked because they are not directly available, or because other more salient aspects attract attention, can be described by the information processing profile. A good example is the different dialogue or interface styles, that also were the topic of chapter 7. When comparing a command interface with a direct manipulation graphical interface, there are several aspects of task performance that are salient. To begin with, the apparent intuitivity of the direct manipulation style is observed directly. Secondly, the ease of use, i.e. the minimal learning necessary to work with the direct manipulation interface is a salient detail. For non-experienced mouse-users the difficulty in using the mouse will be observable. These aspects are the most important and salient differences between these two interfaces. Other differences that are less easily observed or do not automatically attract attention, tend to be overlooked. The information processing during task execution in a command interface differs markedly from that using a direct manipulation graphical interface. The latter interface requires more perceptual and motor activity while the former requires more cognitive activity. These differences seem less important than intuitiveness or learnability, but in certain situations they may be of critical importance.

The detailed information processing analysis is especially important for task situations in which people perform the same task for many hours a day. An example is the work of people typing ZIP-codes in a semi-automatic postal processing setup. These people perform their routine task for 8 hours a day, 5 days a week. In that case, the detailed information processing profile can be of great use. Within such situations, specific fatiguing effects, such as visual fatigue, are likely to occur, and possibly can be prevented or anticipated with the help of an information processing profile.

Another situation in which the information processing profile will be important, is when fast reactions are required or where potentially severe dangers are involved, such as in flying aircrafts, controlling nuclear power plants or air traffic control. In such situations, small differences in the level of (specific) fatigue, small differences in the information processing or small differences in reaction time, can be the difference between a catastrophy or smooth performance.

Especially in the latter example, an average information processing profile will not suffice. A dynamic representation of information processing at any moment in the task is needed. The average profile can obscure differences between different parts of the task (see also Neerincx, 1995; Neerincx et al. 1998).

The information processing profile should be seen as a first step towards an information processing analysis. It has been shown in this thesis that the approach is valid. Next, the approach should be further developed into a more dynamic representation of information processing. In addition, it should be assessed whether the information processing model is detailed enough or that perhaps a more detailed architecture like the EPIC architecture is needed (Meyer & Kieras, 1994). An advantage of the more elaborate EPIC architecture is that it differentiates between visual, auditive and tactile processing and between vocal, manual and oculomotor behavior. This could proof to be a valid distinction in relation to the usefulness of the information processing profile (see also Kieras, 1999).

8.1.3 Constraints of the NGOMSL-IPA and future developments

In chapter 2 the advantages of the GOMS approach were already mentioned. In addition, the extensions to the NGOMSL approach that were discussed in this thesis have made a relationship with information processing, allowing for the estimation of mental workload. This adds to the list of advantages of the GOMS approach. Nevertheless, GOMS (and NGOMSL-IPA) also suffer from several constraints.

Having chosen a GOMS-like approach means having chosen a detailed and low-level analysis. Although the GOMS-approach is not as detailed as the Interacting Cognitive Subsystems approach (Barnard, 1987), it is substantially more detailed than the Cognitive Task Load approach (Neerincx, 1995; 1999; Neerincx et al. 1998). This means that the applicability of the approach will be limited. A very detailed approach necessarily means that performing a task analysis is time consuming and requires a relatively high degree of expertise within the domain of task modelling and cognitive science. As a result, it is not easily applicable by human-computer interface designers without any specialised training. This contrasts with the Cognitive Task Load approach that is designed to be easily usable by software designers and is relatively simple and moderately time-consuming. The choice for a detailed and relatively complex approach like the GOMS approach is a deliberate decision. Real-life tasks usually are very complex. It would be a misunderstanding to think that such complex tasks can be analysed in a short time, by an analyst that has virtually no specific training in and knowledge of cognitive science, and can be described in a simple model. If a task is complex, a model describing it necessarily is complex as well. Especially in the case of computer supported tasks that require a lot of mental processing, a task model must reflect that mental character and cannot reduce it to a few simple mental operations. A simple model would not do justice to the complexity of the task and is therefore bound to be of limited use or no use at all. In addition, it would not be a good thing to let the choice for a method be dependent on the level of skill and knowledge of the average user-interface designer. Performing a cognitive task analysis should not be done by someone without knowledge of cognitive psychology. Either designers should be trained in cognitive science, or a cognitive scientist should perform the analysis, if the task model has to be valid and relevant.

Often, an anlysis of only a part of a task is sufficient. Many tasks consist of repetitive actions, e.g. menu selection in a computer application. In this case, modelling only a part of a task, or one task instance suffices. In addition, design decisions often concern specific components of an interface and thus specific parts of a task. In those cases an analysis of the task part involved suffices. The complexity of the NGOMSL-IPA approach is therefore often manageable.

There are some other limitations to the NGOMSL-IPA approach that limit its applicability or the scope of its conclusions. Task performance is viewed as goal oriented behavior. Task behavior is modelled as if completely driven from within the human operator. This is a limited view on task performance which misses the richness of the interaction with the environment (Suchman, 1987; 1993; Vera & Simon, 1993; Norman, 1993). This can partly be solved within an NGOMSL-IPA model with the inclusion of operators that are dependent on incoming information and by providing several different task-routes and strategies, dependent on information within the environment, through decision operators and selection rule sets. It is not possible, however, to predict what can happen in a complex environment and what the consecutive actions of the human operator will be.

Another factor that limits the application and scope of the NGOMSL-IPA model is that it can in practice only be used to model expert behavior (just as the original GOMS approach, see Olson & Olson, 1990). The model applies to skilled users, and not to beginners or intermediates. Such non-skilled users spend a considerable amount of time engaged in problem-solving behavior, rather than in simply retrieving and executing plans and procedures from memory. The model only describes the plans and procedures, but does not describe performance in a situation where the human operator does not know which actions will solve the problem and has to think of a plan and compile the procedures. Many real-life tasks for which a cognitive task analysis will be performed, are performed by task experts, and in those cases this limitation is not very severe.

A related limitation of the NGOMSL-IPA is that it is well suited for relatively simple tasks that have only a limited degrees of freedom for operator performance. Text-editing tasks are a good example and have been extensively modelled with the GOMS-like analyses. In these tasks, there are not many different ways to perform subtasks, and there is only a limited degree of interaction with the environment. Other tasks, like a process-control task, have many more degrees of freedom. In such tasks there are many ways to reach a goal, there is a rich interaction with the environment and the order in which several actions can be performed is not fixed. This leads to much more complex models and sometimes makes it impossible to make a reliable model, as we have experienced in modelling process-control tasks. This is something that will require further research, but is not principally impossible within the NGOMSL-IPA approach. Even more complex tasks, like tasks which require creative skills (creating/writing text) cannot be analyzed by an approach as the NGOMSL-IPA. These tasks only minimally require procedural knowledge and cannot be described by fixed procedures and actions on a low level. The scope of the analysis that will be discussed in this thesis has been limited to relatively simple and well-ordered tasks. Tasks that require more problem solving behavior and that can be described as highly knowledge based can be best analysed by an approach like that of Neerincx (1995; 1999; Neerincx et al. 1998) relating to the theory of Rasmussen (1983).

In general, within a GOMS model, errors are usually not modelled (Olson & Olson, 1990).

Partly because this is very difficult, because some errors cannot be predicted and henceforth not be modelled, partly because it would lead to very large and complex models. Some errors (execution errors, forgetting errors) could be included in the GOMS approach, some not (errors because of misunderstanding the task or the system). Within the approach in this thesis, errors have not been included in modelling.

There are more limitations to the original GOMS approach that have been solved by other GOMS-derivatives (Olson & Olson, 1990). The NGOMSL-IPA approach is a first step in solving some other limitations of the original GOMS approach (Olson & Olson, 1990) regarding mental workload and information processing. Modelling learning and transfer of knowledge has been done by Polson & Kieras (1985; Kieras & Polson, 1985). Modelling error behavior has been done by Lerch and coworkers (Lerch, Mantei & Olson, 1989). Parallel processing has been modelled by John (1988). John and coworkers also extended the GOMS model to more complex task domains that require a high level of interaction with the environment (computer games, browser tasks) and that have many degrees of freedom (John, Vera & Newell, 1990; Peck & John, 1992), using the SOAR architecture which can also include problem solving behavior. Karwowski, Kosiba, Benabdallah and Salvendy (1990) extended the GOMS model to a fuzzy model that makes the model more valid and in accordance with real task behavior, as well as better applicable in complex task environments. Kieras & Meyer (1997) and Meyer and Kieras (1999) have further developed the Model Human Processor into a more sophisticated architecture called EPIC. In addition, Kieras (1999) has developed a GOMS model simulation tool (GLEAN3) to assist in making a task model, which is based on the EPIC architecture.

The research discussed in this thesis has shown that the NGOMSL-IPA approach is valid and is therefore worthwhile to develop further. These developments can go in any direction, but some useful extensions will shortly be mentioned.

Perhaps the most serious limitation of the NGOMSL-IPA is that it cannot model problem solving behavior. This could be tackled in the way Young & Whittington (1990) have shown. They extended GOMS modelling to tasks that involve many problem solving characteristics. In essence it means that a GOMS model is constructed which contains several blanks. These blanks are parts of a task that cannot be completely modelled. The next step would be try to fill in these blanks, e.g. by stating the various possible ways that part of the task could be performed and indicating which factors influence the actual choice of strategy. This has already been done in our laboratory by using the NGOMSL-IPA to model an information search in a large data-base. Other steps still remain to be made, however. The inclusion of SOAR (John, Vera & Newell, 1990) or ACT-R (Anderson, 1993; Anderson & Lebiere, 1998) probably could partly solve the problem. John and coworkers (John et al., 1990; Peck & John, 1992) have shown that using SOAR enables the modelling of more complex tasks such as browser tasks or video games, that are more problem solving in character. The inclusion of fuzzy techniques (Karwowski et al., 1990) is also a useful extension in this regard.

A second extension of the NGOMSL-IPA approach could be to make a data-base of tasks and user groups, consisting both of models of standard simple tasks and of complex real-life task models, that could serve as reference points for any new task model. More generally, the approach should be further developed for an application in designing user-interfaces and using it to adjusting tasks to human capacities (e.g. designing for special groups such as elderly or handicapped people; Sikken, Engelmoer & Brouwer, 1994).

To make the approach really usable, an expert system should be developed that can serve as a modelling aid. This expert system should be used as a simple computer tool to diminish the laborious work of writing out an NGOMSL-IPA model and of calculating the cognitive complexity, execution time, learning time, workload and information processing profile. Also it should be used to simulate a model, in order to test it. In addition, the expert system should provide 'intelligent' help for performing an NGOMSL-IPA task analysis.

8.2 Human-computer interaction

8.2.1 Applying NGOMSL-IPA in human-computer interaction design

GOMS-like approaches are useful for designing human-computer interfaces, or more generally, human-machine interfaces, as has been reported in several studies (e.g. Eberts, 1994; de Vries & Johnson, 1992; John, 1988, Gugerty, Halgren, Gosbee & Rudisill, 1991). Especially the predictive character of keystroke level models is essential in this regard. The designer of an interface can use GOMS-like models, including NGOMSL-IPA models, to compare different design alternatives.

The NGOMSL-IPA approach can be used to compare several alternative interface designs. Tasks performed within all the alternatives can be analyzed and for each interface the (partly) quantitative estimates can be calculated. Subsequently, the time to perform the task, the load on working memory, the information processing profile, the ease of learning, the complexity and the mental workload can be used to choose between the interfaces.

The exact criterion for selecting a design is situation and task dependent. In designing an interface for combat aircraft control, reaction time will be a critical factor and learning time will be relatively unimportant. For the design of an interface of a CAD-CAM program, complexity, error-proneness and balanced use of the processors generally will be important aspects, while reaction time will be relatively unimportant. The fact that the NGOMSL-IPA approach allows the use of several distinct estimates, makes it a powerful tool with which many quite different tasks and situations can be tackled. Using the NGOMSL-IPA approach can lead to designing better user-interfaces, which can be based on the explicitly stated task structure.

The NGOMSL-IPA approach forces the analyst to think about the cognitive aspects of task performance and look beyond the overt actions. It extends the theorizing about task performance to the cognitive domain. Other, equally important aspects, such as memory load or perceptual processing, can in this way also dictate the design of user-interfaces.

It is difficult to relate (partly) quantitative estimates to an absolute standard. The speed of task execution can be compared between two alternative designs, but it cannot be said if it is fast or slow altogether. The same applies to the workload involved in a task. The method could therefore be used to choose the best of a few alternatives, but does not provide an absolute reference point.

It is important that the estimates should be related to individual users, and if this is not possible, to specific groups of users. Card et al. (1983) introduced the slow man, middle man and fast man estimates, and the results from e.g. chapter 4 clearly showed the individual differences that exist between estimates.

A special application of GOMS techniques is in the design of help and documentation (Elkerton & Palmiter, 1990; Gong & Elkerton, 1990). Critical in the design of help and documentation is the link between the user and the information database. In order to optimize this link, the help information should be presented in a goal oriented structure, such that users can directly find the information needed to attain their goals. Traditional help systems are structured by the program or interface, while they should be structured by the task (goals and subgoals) of the users. The log-file analysis (chapter 3) can be specifically useful in shaping on-line help. From the contents of a log, the actions of the user, in relation to the task he/she performs, his/her goals and subgoals can be inferred. When this is done on-line, the user can be presented with on-line help that is related to and described in terms of the goals he/she is pursuing. The design of such a help system requires a detailed task analysis. A step further would be to also include the depth of the goal structure and psychophysiological indices in order to detect occurrences of high mental workload and take this into account in presenting the help (or other information) to the user. This specific set-up would only be useful in tasks and situations that are time critical or carry large potential danger.

The NGOMSL-IPA approach is a very detailed approach. The best use in design is probably a kind of top-down approach, starting out with high level analysis (possibly also a GOMS-like analysis, but not necessarily), and only pursuing the analysis to a very detailed level for those parts of a task, where there is a clear question regarding one of the estimates or where something can be gained by the specifics of the NGOMSL-IPA.

8.2.2 Towards a psychophysiology of human-computer interaction

The experiments decribed in this thesis have made extensive use of psychophysiology. Several indices that can be calculated from heart rate and the EEG have been described and shown to be applicable in the context of human-computer interaction. Specifically, they were used to evidence the operations of the perceptual processor, the cognitive processor and the motor processor, and to estimate the amount of workload involved in task performance. In this paragraph some general observations from the use of psychophysiological measures in the text editing tasks, will be made.

8.2.2.1 Integrating measures

The three text-editing experiments clearly showed the usefulness of integrating behavior and psychophysiological indices. The questions pursued in these experiments could not have been answered by performance measures alone, nor by psychophysiological measures alone. Psychophysiological indices served a double function: they were used to estimate the costs of performance and they were used to make covert perceptual and cognitive processing visible. The experiments clearly showed the need for a psychophysiology of human-computer interaction, since that is the only way to clarify the mental processing that is so abundant in computer supported tasks. The argument had already been put forward by others (Gale & Christie, 1987; Wastell, 1990; Wiethof et al., 1991; Wiethof, 1997; van Westrenen, 1999), but has been empirically supported by the text editing data presented in chapter 6 and 7.

There are some practical difficulties, hindering a wide-spread use of psychophysiology in human-computer interaction. Expensive and sophisticated equipment is needed, special skills are required for the experimenters, time consuming procedures are needed for recording and analysis, and the subjects will at least experience the situation somewhat awkward. Although these difficulties are only minor if only heart rate variability is used, still they will dictate a special attitude towards psychophysiological measures in applied situations. Like the application of a very detailed cognitive task analysis, psychophysiological measures should only be applied if a special question is asked or if special information is required (see e.g. Byrne & Parasuraman, 1996). Still, using heart rate variability can easily be implemented as a standard procedure in any human-computer interaction research. An essential pre-requisite for a usable methodology is that the analysis must be performed in a standardized way in batch-mode, because large quantities of data are required and many repeated measures have to be performed.

The various performance, self-report and psychophysiological indices all reflect different aspects of task performance and mental load. The individual indices often are not convincing enough, and could be subject to multiple interpretations. Yet, if several, in themselves non-conclusive indices, all point in the same direction, then this aggregation provides strong converging evidence.

Some self-report scales also measure invested effort or mental workload, just as some psychophysiological indices. Nevertheless, it is wise to include both in an experimental investigation, because they appear sensitive to different aspects of mental workload. This was evidenced in chapter 7, where the SMEQ and the NASA-TLX did not mirror heart rate variability in the mid-frequency band, while all seem to be sensitive to mental workload.

Applying psychophysiology in human-computer interaction requires a special experimental set-up, e.g. a set-up as was described in chapter 3. It is essential that the elementary actions are recorded in real time with a high time resolution. These should be related to psychophysiological variables in a meaningful manner, i.e. behavior should be interpreted before integration with psychophysiology. A general approach as was used in the research from this thesis in itself would be best, but would be hard to implement on some computer systems. Especially direct manipulation graphical systems require only few different motor actions (mouse moves and clicks), whose meaning depends on the location of the mouse cursor and the actual screen configuration. In that case, it will be difficult to translate mouse actions into their actual meaning within the task. In addition, some computer systems or computer operating systems do not support real-time programming. The set-up thus has its limitations.

8.2.2.2 The probe-evoked potential

The use of irrelevant stimuli for the analysis of probe evoked potentials was put forward as a special kind of dual task methodology. The irrelevant probes do not disturb tak performance as does the inclusion of a regular second task. They provide a nice opportunity to non-invasively measure spare capacity.

The morphology of the probe-evoked potential shows a P1, N1 and P2 complex, of which the P1 is rather small. The N1 and P2 have a fronto-central maximum. Both the N1 and the P2 were shown to be sensitive to changes in workload in previous studies reported in the literature. In chapter 5 it was shown that the probe-evoked potential is sensitive to differences in working memory load. The findings were comparable to those from the literature: the N1 is enlarged with an increase in workload, while the P2 decreases with increasing workload.

Other studies in the literature have reported a P300 component, which also is sensitive to workload differences (Trejo, Lewis & Blankenship, 1987; Sirevaag, Kramer, Wickens, Reisweber, Strayer & Grenell, 1993). As was argued in chapter 5, the supposedly P300 component should be called a P2 peak. There are several reasons for that. To begin with, P300 components are usually generated by task relevant stimuli, but not by irrelevant stimuli (Sutton & Ruchkin, 1984). Kramer, Trejo & Humphrey (1995) also recognized that point and also reported no P300 peak. Secondly, the so-called P300 is fronto-centrally maximal (Trejo et al., 1987; Sirevaag et al., 1993), while the P300 has a parietal maximum (although the P3a has a more frontal maximum). The latency of the peak is around 200 ms poststimulus, which is very early for a P300. Thus, what is sometimes called the P300 in probe-evoked potential studies, is the same component as was described as the P2 in this thesis. The P300 results reported match those reported on the P2 component, supporting the argumentation that P2 is a better indication.

The studies in the literature using the probe-evoked potential technique have led to a variety of results, which are sometimes contradictory. The most consistent result is a smaller amplitude in the P2 range with increasing workload (Bauer et al., 1987; Trejo et al., 1987; Sirevaag et al. 1993) as was also witnessed in the memory search experiment from chapter 5 and the text editing experiments from chapter 6. The second effect that is regularly reported is an effect on the N1, which is larger (more negative) for higher workload conditions (Näätänen, 1975; Papanicolaou et al., 1984), as was also evidenced in the memory search task from chapter 5. Occasionally, an effect the other way around, a smaller N1 with higher workload, is also reported (Kramer et al., 1995).

Papanicolaou et al. (1984) already presented a thorough review of the probe-evoked potential literature, and described many methodological problems that have led to high variability between studies and led to difficulties in interpretation. Some will be discussed again here, and some new issues will be introduced.

Some authors have used a base-line condition as a reference for task probe-evoked potentials (Kramer et al., 1995; Papanicolaou, 1984). In the Kramer et al. (1995) study, the auditory N1 was larger in this base-line condition than in the task condition, which is not in accordance with the data from chapter 5. Such a base-line condition is taken as an episode of very low workload, while actually the workload cannot be unequivocally be assessed. The instructions to the subjects in such a base-line condition vary, from passive viewing the screen to reacting to infrequent stimuli. Consequently the workload also varies. In addition, it cannot be assessed what subject actually do when instructed to do nothing. Possibly the subjects do nothing overtly, but are heavily engaged in covert activity, e.g. preparation for the task. The use of a base-line condition can possibly introduce more noise, instead of clarity.

In the tradition of the oddball paradigm several studies have used frequent and infrequent probe-stimuli (Trejo et al, 1995; Kramer et al., 1995; memory search study from chapter 5 and experiment 2 from chapter 6). Although the approach is very appealing, it has been proven to be very difficult to apply, because there have to be at least four times as many frequent probes as infrequent probes. This means that under normal task conditions, with a moderate stimulus rate, it is very difficult to get enough deviant stimuli to calculate a reliable probe-evoked potential. Up to now it has not been very succesful.

The frequent-infrequent distinction has a long tradition in auditory oddball tasks, and can easily be applied in auditory probes. Applying it with visual probes, i.e. by making color the attribute that differs between frequent and infrequent probes, has not been shown to be succesful, possibly owing to problems in the signal to noise ratio (Trejo et al., 1995).

As was evident in the experiments in this thesis, there is an important difference between visual and auditive probe-evoked potentials. Both showed comparable effects in the memory search task, although the N1 of the stimuli presented shortly after the display set, indicated some modality specific effect. Next, the second text editing experiment showed an absence of any effect on the auditory probes while there was an effect on the visual probes, again indicating some modality specific influence. Kramer et al. (1995) also failed to find an effect